

多齿红山茶叶绿体基因组序列特征分析

童一涵, 郑倩, 杜新明, 冯士令, 周莉君, 丁春邦, 陈涛^①

(四川农业大学生命科学院, 四川雅安 625014)

摘要: 本研究利用高通量测序技术对多齿红山茶 (*Camellia polyodonta* How ex Hu) 叶片进行核基因组重测序, 并对组装得到叶绿体基因组进行注释、序列特征和系统发育分析。结果显示: 多齿红山茶叶绿体基因组长度为 156 778 bp, 由 4 部分组成, 其中, 大单拷贝区长度为 86 493 bp, 小单拷贝区长度为 18 219 bp, 2 个反向重复区长度均为 26 033 bp。注释到 87 个 CDS 基因、37 个 tRNA 基因和 8 个 rRNA 基因, 分为光合作用基因、自我复制基因、其他基因和未知功能基因 4 类。多齿红山茶叶绿体基因组中共检测到 52 个 SSR 位点, 且均为单碱基 (A/T) 的重复类型。多齿红山茶叶绿体基因组相对同义密码子使用度大于 1.00 的密码子共有 30 个 (终止密码子除外), 其中, 有 27 个以 A 或 U 结尾, 有 3 个以 C 或 G 结尾。山茶属 (*Camellia* Linn.) 植物 *rps19*、*rpl2* 和 *trnH* 基因相对保守, 长度分别为 279、1 495 和 75 bp, 与反向重复区边界间的相对位置也保持一致, 而 *ndhF*、*ycf1* 和 *trnN* 基因在种间具有特异性; 与木荷 (*Schima superba* Gardn. et Champ.) 相比, *rps19*、*rpl2* 和 *trnH* 基因位置和长度在属间差异明显, *ndhF*、*ycf1* 和 *trnN* 基因位置和长度在属间差异更明显。基因不同程度的扩张和收缩是造成反向重复区和小单拷贝区长度差异的原因。聚类分析结果显示: 多齿红山茶、滇山茶 (*C. reticulata* Lindl.) 和南山茶 (*C. semiserrata* Chi) 聚为一支, 短柱茶 [*C. brevistyla* (Hayata) Coh. St.]、小果油茶 (*C. meiocarpa* Hu) 和茶梅 (*C. sasanqua* Thunb.) 聚为一支, 聚在一支的种类亲缘关系较近; 红山茶组 (Sect. *Camellia* (Linn.) Dyer)、油茶组 (Sect. *Oleifera* H. T. Chang) 和短柱茶组 (Sect. *Paracamellia* Sealy) 的种类均聚在了不同的分支。综合研究结果表明: 多齿红山茶叶绿体基因组呈典型的四分体结构, 密码子偏好以 A 或 U 结尾, 且高频率使用 AGA 密码子编码 Arg; 山茶属不同物种间 *ndhF*、*ycf1* 和 *trnN* 基因的扩张和收缩情况差异明显; 系统发育分析结果支持将短柱茶组合并入油茶组。

关键词: 多齿红山茶; 叶绿体基因组; 序列特征; 系统发育分析

中图分类号: Q943.2; S685.14; S718.46 文献标志码: A 文章编号: 1674-7895(2022)05-0027-10

DOI: 10.3969/j.issn.1674-7895.2022.05.04

Analysis on sequence characteristics of chloroplast genome of *Camellia polyodonta* TONG Yihan, ZHENG Qian, DU Xinming, FENG Shiling, ZHOU Lijun, DING Chunbang, CHEN Tao^① (College of Life Science, Sichuan Agricultural University, Ya'an 625014, China), *J. Plant Resour. & Environ.*, 2022, 31(5): 27-36

Abstract: The nuclear genome of *Camellia polyodonta* How ex Hu leaf was re-sequenced by using high-throughput sequencing technology, and annotation, sequence characteristics and phylogenetic analysis were conducted for the assembled chloroplast genome. The results show that the length of chloroplast genome of *C. polyodonta* is 156 778 bp, which is composed of four regions, in which, the length of large single copy region is 86 493 bp, that of small single copy region is 18 219 bp, and the lengths of two inverted repeat regions are both 26 033 bp. 87 CDS genes, 37 tRNA genes, and 8 rRNA genes are annotated, which can be divided into 4 categories of photosynthesis genes, self-replication genes, other genes and unknown function genes. 52 SSR loci are detected in chloroplast genome of *C. polyodonta*, and

收稿日期: 2021-12-17

基金项目: 国家自然科学基金资助项目 (31801826); 四川省科技计划项目 (2020YJ0143); 雅安市科技项目 (21SXHZ0017)

作者简介: 童一涵 (2001—), 男, 湖北武汉人, 本科, 主要从事植物种质资源与系统发育方面的研究。

^①通信作者 E-mail: chentao293@163.com

引用格式: 童一涵, 郑倩, 杜新明, 等. 多齿红山茶叶绿体基因组序列特征分析[J]. 植物资源与环境学报, 2022, 31(5): 27-36.

all of them are of single base (A/T) repeat type. There are 30 codons (except for stop codons) with the relative synonymous codon usage greater than 1.00 in chloroplast genome of *C. polyodonta*, in which, 27 codons are ended with A or U, 3 codons are ended with C or G. *rps19*, *rpl2*, and *trnH* genes of *Camellia* Linn. species are relatively conserved, their lengths are 279, 1 495, and 75 bp respectively, their relative locations with respect to boundaries of inverted repeat regions are also consistent, while *ndhF*, *ycf1*, and *trnN* genes are specific among species; compared with *Schima superba* Gardn. et Champ., the positions and lengths of *rps19*, *rpl2* and *trnH* genes are significantly different among genera, while those of *ndhF*, *ycf1* and *trnN* genes are more significantly different among genera. Different degrees of gene expansion and contraction are responsible for the differences in lengths of inverted repeat and small single copy regions. The clustering analysis result shows that *C. polyodonta*, *C. reticulata* Lindl. and *C. semiserrata* Chi are grouped into one branch, the species clustered in one clade are closely related; Sect. *Camellia* (Linn.) Dyer, Sect. *Oleifera* H. T. Chang and Sect. *Paracamellia* Sealy are all clustered in different branches, and *C. brevistyla* (Hayata) Coh. St., *C. meiocarpa* Hu and *C. sasanqua* Thunb. are grouped into one branch. The comprehensive analysis result shows that the chloroplast genome of *C. polyodonta* shows a typical tetrad structure, the codons prefer to end with A or U, and AGA codon is frequently used to code Arg; the differences in expansion and contraction of *ndhF*, *ycf1* and *trnN* genes among different species in *Camellia* are evident; the phylogenetic analysis result supports that incorporation of Sect. *Paracamellia* Sealy into Sect. *Oleifera* H. T. Chang.

Key words: *Camellia polyodonta* How ex Hu; chloroplast genome; sequence characteristics; phylogenetic analysis

多齿红山茶(*Camellia polyodonta* How ex Hu)隶属于山茶科(Theaceae)山茶属(*Camellia* Linn.),又称宛田红花油茶,1965年由How ex Hu定名并且将其归入山茶属红山茶组[Sect. *Camellia* (Linn.) Dyer]^{[1]58}。红山茶组植物花色鲜红,花期长,雨后花色不变,被广泛应用于园林城市、旅游景区等的绿化,是西部地区发展观光旅游的优良树种^[2]。中国是红山茶组植物的分布中心,种质资源丰富,其中多齿红山茶主要分布在海拔900 m的山坡阔叶林和林缘。由于红山茶组植物所处的自然地理环境不同,物种广泛重叠和自然杂交,导致种间变异,特别是具有栽培历史的观赏花卉和油料植物,在进一步人工引种栽培后,还会引起染色体多倍化,使种间变异更为复杂多样,导致红山茶组植物的种类划分、扩散路线、分化和进化趋势尚存在较大分歧。

许多专家从分类学角度对山茶属植物进行了相关的研究,目前,国内主要参考张宏达^{[1]5}和闵天禄^[3]的分类系统,但这2个分类系统之间存在很大的差异,其原因为形态学的传统物种分类方法易受环境因子影响^[4]。而以分子钟理论为基础,精确地研究物种之间的系统发育关系,可为解决存在争议的分类问题提供更加可靠的理论依据和数据参考^[5]。20世纪60年代初,在地钱(*Marchantia polymorpha* Linn.)^[6]中首次报道了叶绿体基因组全序列。植物叶绿体基

因组远小于核基因组,具有分子量适中、便于测序、多拷贝、结构简单、DNA的核苷酸置换率适中、编码区和非编码区的分子进化速度差异显著以及各类群叶绿体基因组之间具有良好的共线性的特点^[7],且植物叶绿体具有一套独立于核基因组外的母系遗传的叶绿体基因组^[8]。随着系统发育学和基因组学的交融,在植物系统发育研究中,基于叶绿体基因组的系统发育基因组学研究优势渐显端倪^[9],为一些分类困难类群的系统学问题提供了新的解决方案。

目前,多齿红山茶叶绿体基因组信息缺乏,关于山茶属进化分类的相关研究结论的可靠性有限,影响了该优良资源的进一步开发和利用,随着第二代高通量测序技术的发展,植物全基因组测序具备了速度快、通量高、成本低、高精度的特点^[10]。鉴于此,本研究使用Illumina测序平台对未被GenBank收录的多齿红山茶进行高深度重测序,利用测得的核基因组序列信息组装出叶绿体基因组,对其序列特征及系统发育进行分析,以期多齿红山茶资源的开发和利用、近缘种间的系统发育关系及山茶属植物的进化和分类研究提供理论依据和参考。

1 材料和方法

1.1 材料

供试材料多齿红山茶为四川省雅安市天全县城

厢镇大岗山(东经 102°46'59"、北纬 30°04'25")原生分布单株,树龄约 20 a,于 2021 年 3 月每株采集新稍嫩叶 3~5 枚,共采集 6 株,混合后装入含有硅胶的自封袋带回实验室。

1.2 方法

1.2.1 总 DNA 提取及测序 使用改良 CTAB 法^[11]提取叶片总 DNA,用质量体积分数 1%琼脂糖凝胶电泳检测 DNA 的完整性,使用 NanoDrop 2000 分光光度计(美国 Thermo Scientific 公司)检测 DNA 浓度和纯度,将总 DNA 进行随机打断、末端修复、连接接头构建 500 bp 的测序文库,使用 HiSeq 2500 高通量测序平台(美国 Illumina 公司)进行双末端重测序,测序由深圳华大基因股份有限公司完成。原始数据剔除接头污染和低质量序列,得到 3.92 Gb 合格数据。

1.2.2 叶绿体基因组组装与注释 过滤后的数据使用 GetOrganelle 软件^[12]进行多齿红山茶叶绿体基因组的从头组装,最后获得 1 条环形的叶绿体基因组序列。使用 CPGAVAS2 在线注释工具(<http://47.90.241.85:16019/analyzer/home>)^[13]对获得的叶绿体基因组序列信息进行在线注释,以 NCBI 上已发表的近缘种滇山茶(*Camellia reticulata* Lindl.)(GenBank 登录号 KJ806278.1)的叶绿体基因组信息作为参考序列,其他参数设为默认值,经人工校正后,将多齿红山茶叶绿体基因组序列上传至 NCBI(GenBank 登录号 OK377261),最后再使用 Chloroplast 在线工具(<https://irscope.shinyapps.io/Chloroplast/>)^[14]绘制叶绿体基因组图谱。

通过 JSHYCloud 在线工具集(<http://cloud.genepioneer.com:9929>)分析并统计叶绿体基因组、大单拷贝区(LSC)、小单拷贝区(SSC)和反向重复区(IR)的 GC 含量。

1.2.3 重复序列与 IR 区边界结构差异分析 使用 MISA 在线工具(<https://webblast.ipk-gatersleben.de/misa/index.php>)^[15]的微卫星定位检测技术对多齿红山茶叶绿体基因组序列中的简单重复序列进行搜索,参数设置参考文献[16],单核苷酸、二核苷酸、三核苷酸、四核苷酸、五核苷酸、六核苷酸的最小重复值分别设置为 10、6、5、5、5、5。

通过 JSHYCloud 在线工具集分析叶绿体基因组 IR 区边界结构差异,从 NCBI 中选同为红山茶组的南山茶(*C. semiserrata* C. W. Chi)(GenBank 登录号 MZ403753.1)和毛蕊红山茶[*C. mairei* (Lévl.)

Melch.](GenBank 登录号 KY406767.1)叶绿体基因组来比对同组不同物种间叶绿体基因组异同,选油茶(*C. oleifera* Abel.)(GenBank 登录号 MN078090.1)来比对不同组间物种的叶绿体基因组异同,选择与山茶属亲缘关系最近的木荷属(*Schima* Reinw.)的木荷(*S. superba* Gardn. et Champ.)(GenBank 登录号 MH782179.1)作为外类群。

1.2.4 密码子偏好性分析 通过 Pasteur Galaxy 在线工具集(<https://galaxy.pasteur.fr/CodonW>)^[17]中的 CodonW 模块分析密码子使用情况,设置输出结果为有效密码子数(ENC)和相对同义密码子使用度(RSCU),其他参数设为默认值^[18]。ENC>35 表示含有较多种类的稀有密码子,且基因表达量偏低;某一密码子的 RSCU>1.00 表示编码对应的蛋白质时偏好使用该密码子,RSCU<1.00 表示不偏好使用该密码子,RSCU=1.00 表示该密码子没有偏好性。

1.2.5 系统发育分析 将多齿红山茶叶绿体基因组序列上传至 NCBI 进行 BLASTn 比对,选择 highly similar sequence (megablast)来比较相似性在 95%以上的序列,检索获得多齿红山茶的近缘种,以明确多齿红山茶叶绿体基因组序列在山茶属中的系统关系。

从比对结果中筛除栽培种和地方种,下载山茶属 22 个野生近缘种,使用 MAFFT v7 软件^[19]进行多序列比对后,使用 MEGA X 软件^[20]校正序列,使用 ModelFinder 软件^[21]计算出最佳的极大似然法(ML)建树模型,再以木荷叶绿体基因组序列作为外类群,使用 IQ-TREE 软件^[22]建树,设置自展支持率为 1 000,其他参数设为默认值,最后使用 iTOL v5 在线工具(<https://itol.embl.de/>)^[23]和 ChiPlot 网站(<https://www.chiplot.online/>)调整系统发育树。

2 结果和分析

2.1 多齿红山茶叶绿体基因组序列特征分析

经过测序组装的完整的多齿红山茶叶绿体基因组长度为 156 778 bp,基因组图谱(图 1)显示:叶绿体基因组呈典型的四分体结构,由 1 个大单拷贝区(LSC)、1 个小单拷贝区(SSC)和 2 个反向重复区(IR)共 4 部分组成,其中,LSC、SSC 和 IR 区的长度分别为 86 493、18 219 和 26 033 bp。基因组的总 GC 含量为 37.33%,其中,LSC、SSC 和 IR 区的 GC 含量分别为 35.34%、30.60%和 42.98%。

成的基因有 19 个, 包括 11 个 CDS 基因和 8 个 tRNA 基因, 其中有 4 个基因在 IR 区重复; 由 2 个内含子和 3 个外显子构成的基因有 *ycf3* 和 *clpP*, 均为 CDS 基因; *rps12* 基因为反式剪切, 只统计 2 个外显子。

表 1 多齿红山茶叶绿体基因列表
Table 1 Gene list of chloroplast genome of *Camellia polyodonta* How ex Hu

基因 ¹⁾ Gene ¹⁾	数量 Number	基因翻译产物 Gene translation product
光合作用基因 Photosynthesis gene		
<i>atpA, atpB, atpE, atpF, atpH, atpI</i>	6	ATP 合成酶 ATP synthase
<i>psaA, psaB, psaC, psaI, psaJ, ycf3</i>	6	光系统 I Photosystem I
<i>psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ</i>	15	光系统 II Photosystem II
<i>ndhA, ndhB * , ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK</i>	12	NADH 脱氢酶 NADH dehydrogenase
<i>petA, petB, petD, petG, petL, petN</i>	6	细胞色素 b/f 复合体 Cytochrome b/f complex
<i>rbcL</i>	1	核酮糖-1,5-二磷酸羧化酶大亚基 Large subunit of rubisco
自我复制基因 Self-replication gene		
<i>rpl14, rpl16, rpl2 * , rpl20, rpl22, rpl23 * , rpl32, rpl33, rpl36</i>	11	核糖体大亚基 Large subunit of ribosome
<i>rpoA, rpoB, rpoC1, rpoC2</i>	4	RNA 聚合酶 RNA polymerase
<i>rps2, rps3, rps4, rps7 * , rps8, rps11, rps12 * , rps14, rps15, rps16, rps18, rps19</i>	14	核糖体小亚基 Small subunit of ribosome
<i>rrn16 * , rrn23 * , rrn4.5 * , rrn5 *</i>	8	rRNA
<i>trnA-UGC * , trnC-GCA, trnD-GUC, trnE-UUC, trnF-GAA, trnG-GCC, trnG-UCC, trnH-GUG, trnI-CAU * , trnI-GAU * , trnK-UUU, trnL-CAA * , trnL-UAA, trnL-UAG, trnM-CAU, trnN-GUU * , trnP-UGG, trnQ-UUG, trnR-ACG * , trnR-UCU, trnS-GCU, trnS-GGA, trnS-UGA, trnT-GGU, trnT-UGU, trnV-GAC * , trnV-UAC, trnW-CCA, trnY-GUA</i>	37	tRNA
其他基因 Other gene		
<i>accD</i>	1	乙酰辅酶 A 羧化酶亚基 Subunit of acetyl-CoA-carboxylase
<i>ccsA</i>	1	细胞色素 C Cytochrom C
<i>cemA</i>	1	囊膜蛋白 Envelop membrane protein
<i>clpP</i>	1	蛋白酶 Protease
<i>infA</i>	1	翻译起始因子 Translational initiation factor
<i>matK</i>	1	成熟酶 Maturase
未知功能基因 Unknown function gene		
<i>ycf1, ycf15 * , ycf2 * , ycf4</i>	6	保守开放阅读框 Conserved open reading frame

¹⁾ * : 表示基因有 2 个 Representing there are two of genes.

表 2 多齿红山茶叶绿体基因组中具有多个外显子的基因信息
Table 2 Information of gene with multiple exons of chloroplast genome of *Camellia polyodonta* How ex Hu

基因 Gene	位置 ¹⁾ Strand ¹⁾	序列长度/bp ²⁾ Length of sequence ²⁾					基因 Gene	位置 ¹⁾ Strand ¹⁾	序列长度/bp Length of sequence		
		外显子 I Exon I	内含子 I Intron I	外显子 II Exon II	内含子 II Intron II	外显子 III Exon III			外显子 I Exon I	内含子 I Intron I	外显子 II Exon II
<i>trnK-UUU</i>	LSC	37	2 493	35			<i>petD</i>	LSC	8	747	475
<i>rps16</i>	LSC	40	855	213			<i>rpl16</i>	LSC	9	1 021	399
<i>trnG-GCC</i>	LSC	23	700	37			<i>rpl2</i>	IR	391	670	434
<i>atpF</i>	LSC	145	716	410			<i>ndhB</i>	IR	775	679	758
<i>rpoC1</i>	LSC	434	741	1 618			<i>trnL-GAU</i>	IR	42	947	35
<i>ycf3</i>	LSC	124	720	230	737	153	<i>trnA-UGC</i>	IR	38	812	35
<i>trnL-UAA</i>	LSC	37	519	50			<i>ndhA</i>	SSC	553	1 085	539
<i>trnV-UAC</i>	LSC	39	587	37			<i>trnA-UGC</i>	IR	38	812	35
<i>rps12</i>	LSC	114	—	243	—		<i>trnL-GAU</i>	IR	42	947	35
<i>clpP</i>	LSC	71	791	291	595	226	<i>ndhB</i>	IR	775	679	758
<i>petB</i>	LSC	6	777	642			<i>rpl2</i>	IR	391	670	434

¹⁾ LSC: 大单拷贝区 Large single copy region; SSC: 小单拷贝区 Small single copy region; IR: 反向重复区 Inverted repeat region.

²⁾ —: *rps12* 基因为反式剪切, 未统计内含子信息 *rps12* gene is trans-clipped, and intron information is not counted.

2.3 多齿红山茶叶绿体基因重复序列分析

多齿红山茶叶绿体基因组中简单重复序列 (SSR) 的类型及分布见表 3。

结果显示:在多齿红山茶叶绿体基因组中共检测到 52 个 SSR 位点,其中有 39 个位于 LSC 区(占比 75.0%),有 9 个位于 SSC 区(占比 17.3%),有 4 个位于 IR 区(占比 7.7%)。这些重复序列均为单碱基

(A/T)的重复类型,其中,重复单元为 A、重复频率为 11 的 SSR 位点数量最多(7),重复单元为 T、重复频率为 10 的 SSR 位点数量最多(12),重复单元为 A/T 的最高重复频率为 17。

2.4 多齿红山茶 IR 区边界结构差异分析

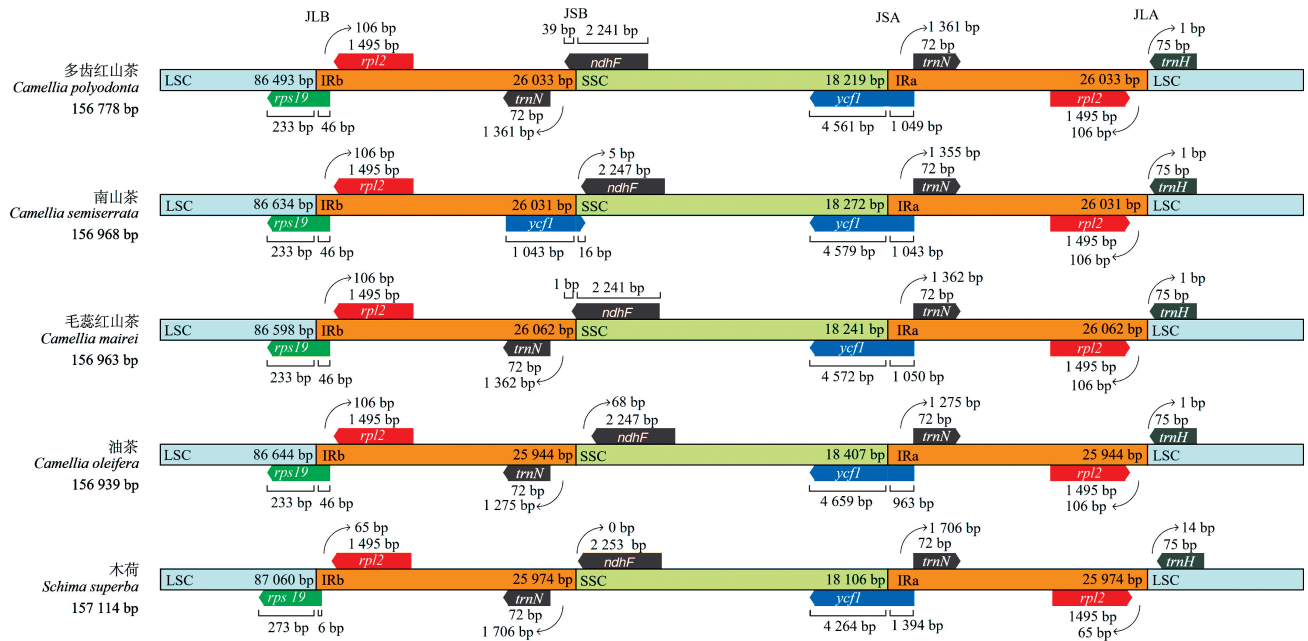
结果(图 2)显示:5 种山茶科植物叶绿体基因组结构从大单拷贝区(LSC)中间呈线性展开,均由 1 个

表 3 多齿红山茶叶绿体基因组中简单重复序列 (SSR) 的类型及分布

Table 3 Type and distribution of simple sequence repeat (SSR) in chloroplast genome of *Camellia polydentata* How ex Hu

重复单元 Repeat unit	重复频率 Repetition frequency	SSR 数量 Number of SSR	起始位点 ¹⁾ Initiation site ¹⁾	重复单元 Repeat unit	重复频率 Repetition frequency	SSR 数量 Number of SSR	起始位点 ¹⁾ Initiation site ¹⁾
A	10	6	3 791, 4 456, 44 098, 65 201, 129 581 * , 138 098 **	T	10	12	27 102, 33 161, 37 138, 56 233, 53 403, 56 233, 65 339, 70 827, 80 495, 82 417, 105 165 ** , 117 804 *
A	11	7	8 852, 17 202, 32 857, 38 178, 83 469, 110 223 ** , 116 977 *	T	11	6	8 709, 14 988, 19 402, 58 882, 84 919, 133 039 **
A	12	3	46 138, 49 253, 129 048 *	T	12	3	13 862, 127 129 * , 128 744 *
A	13	3	356, 37 538, 38 411	T	13	3	15 387, 49 910, 56 667, 115 164 *
A	14	1	46 394	T	14	4	60 660, 65 655, 82 942, 121 176 *
A	15	1	32 520	T	15	1	53 289
A	17	1	12 489	T	17	1	129 962 *

¹⁾ * 和 ** 分别表示分布在小单拷贝区 (SSC) 和反向重复区 (IR) 的单碱基重复序列,无标记表示分布在大单拷贝区 (LSC) 的单碱基重复序列 * and ** represent the single base repeat sequences distributed in the small single copy region (SSC) and inverted repeat region (IR), no label represents the single base repeat sequence distributed in the large single copy region (LSC).



JLB: LSC 和 IRb 区的边界 Boundary of LSC and IRb regions; JSB: IRb 和 SSC 区的边界 Boundary of IRb and SSC regions; JSA: SSC 和 IRa 区的边界 Boundary of SSC and IRa regions; JLA: IRa 和 LSC 区的边界 Boundary of IRa and LSC regions.

图 2 5 种山茶科植物叶绿体基因组大单拷贝区 (LSC)、小单拷贝区 (SSC) 和反向重复区 (IR) 边界位置的比较
Fig. 2 Comparison on the boundary locations of large single copy region (LSC), small single copy region (SSC) and inverted repeat region (IR) in chloroplast genome of five species of Theaceae

LSC 区、1 个小单拷贝区 (SSC) 和 2 个反向重复区 (IR) 4 部分组成, 各个区域间对应的连接基因也基本相同, 仅南山茶注释到了横跨 JSB (IRb 和 SSC 区的边界) 和 JSA (SSC 和 IRa 区的边界) 边界的假基因 *ycf1*。

5 种植物的 *rps19* 基因横跨 JLB (LSC 和 IRb 区的边界) 边界, *rpl2* 基因位于接近 JLB 边界的 IRb 区, *trnN* 基因位于接近 JSB 边界的 IRb 区, 且在接近 JSA 边界的 IRa 区反向重复, 而 *ycf1* 基因横跨 JSA 边界, *rpl2* 基因位于接近 JLA (IRa 和 LSC 区的边界) 边界的 IRa 区, *trnH* 基因位于接近 JLA 边界的 LSC 区。

由 IR 区边界扩张和收缩情况看, 山茶属植物 *rps19*、*rpl2* 和 *trnH* 基因相对保守, 基因位置和长度一致, 且红山茶组与油茶组间无差异; 但与木荷相比, 上述 3 个基因位置和长度在属间差异明显。山茶属植物 *ndhF*、*ycf1* 和 *trnN* 基因则具有位置和长度的特异性, 组间差异明显; 与木荷相比, 属间差异则更加明显, 说明 *ndhF*、*ycf1* 和 *trnN* 基因不同程度的扩张和收缩导致了不同物种间的 IR 和 SSC 区长度差异。

2.5 多齿红山茶密码子偏好性分析

基于多齿红山茶叶绿体基因组中得到的 87 个 CDS 基因序列研究密码子偏好性的常用参数, 经 CodonW 模块分析, 有效密码子数 (ENC) 为 55.3, 明显大于 35, 表明叶绿体基因的表达量偏低, 且基因中还含有较多种类的稀有密码子。

多齿红山茶氨基酸的相对同义密码子使用度 (RSCU) 见表 4。结果显示: RSCU 值大于 1.00 的密码子共有 30 个 (终止密码子除外), 其中, 有 27 个以 A 或 U 结尾, 有 3 个以 C 或 G 结尾, 说明多齿红山茶叶绿体基因组的密码子偏好以 A 或 U 结尾。RSCU 值大于 1.60 的密码子为编码精氨酸 (Arg) 的 AGA, RSCU 值小于 0.60 的密码子包括编码异亮氨酸 (Leu) 的 CUG、编码组氨酸 (His) 的 CAC、编码精氨酸 (Arg) 的 CGC、编码天冬酰胺 (Asn) 的 AAC、编码丙氨酸 (Ala) 的 GCG 和编码天冬氨酸 (Asp) 的 GAC。

综上所述, 多齿红山茶叶绿体基因组高频率使用 AGA 编码 Arg, 低频率使用 CUG、CAC、CGC、AAC、GCG 和 GAC 分别编码 Leu、His、Arg、Asn、Ala 和 Asp。

表 4 多齿红山茶氨基酸的相对同义密码子使用度 (RSCU)

Table 4 Relative synonymous codon usage (RSCU) of amino acids of *Camellia polyodonta* How ex Hu

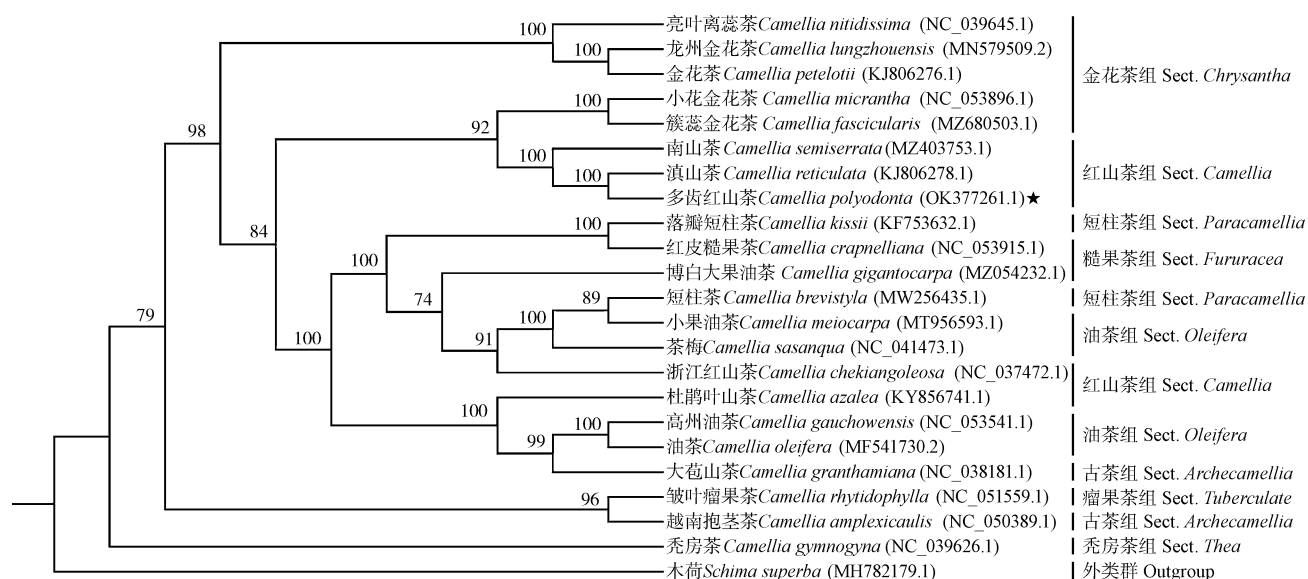
氨基酸 Amino acid	密码子 Codon	数量 Number	RSUC	氨基酸 Amino acid	密码子 Codon	数量 Number	RSUC	氨基酸 Amino acid	密码子 Codon	数量 Number	RSUC
Phe	UUU	2 211	1.17	Arg	CAC	390	0.58	Ala	GUC	413	0.71
	UUC	1 567	0.83		CGU	377	0.69		GUA	730	1.25
Ser	UCU	1 126	1.44		CGC	221	0.41	GUG	414	0.71	
	UCC	877	1.12		CGA	584	1.07	GCU	462	1.18	
	UCA	967	1.23	CGG	357	0.66	GCC	368	0.94		
	UCG	579	0.74	AGA	1136	2.09	GCA	503	1.29		
Leu	AGU	673	0.86	AGG	590	1.08	GCG	231	0.59		
	AGC	478	0.61	Gln	CAA	1 070	1.40	Asp	GAU	1 057	1.44
	UUA	1 183	1.30	CAG	458	0.60	GAC	416	0.56		
	UUG	1 149	1.26	Ile	AUU	1 665	1.16	Glu	GAA	1 358	1.37
	CUU	1 081	1.19	AUC	1 137	0.79	GAG	623	0.63		
	CUC	696	0.77	AUA	1512	1.05	Gly	GGU	562	1.01	
Pro	CUA	851	0.91	Met	AUG	945	1.00	GGC	344	0.62	
	CUG	492	0.54	Thr	ACU	666	1.16	GGA	804	1.44	
	CCU	687	1.12	ACC	571	0.99	GGG	522	0.94		
	CCC	605	0.99	ACA	715	1.24	Tyr	UAU	1 414	1.34	
	CCA	762	1.25	ACG	347	0.60	UAC	701	0.66		
Trp	CCG	393	0.64	Asn	AAU	1 847	1.41	TER * ¹⁾	UAA	1 110	1.14
	UGG	724	1.00	AAC	775	0.59	UAG	809	0.83		
Cys	UGU	661	1.20	Lys	AAA	2 128	1.36	UGA	1 012	1.04	
	UGC	443	0.80	AAG	992	0.64					
His	CAU	946	1.42	Val	GUU	772	1.33				

¹⁾ * : 终止密码子 Stop codon.

2.6 山茶属植物系统发育分析

基于22个山茶属种类和1个木荷属种类的叶绿体基因组构建系统发育树(图3)。结果显示:在山茶属植物中,金花茶[*Camellia petelotii* (Merr.) Sealy]、龙州金花茶(*C. lungzhouensis* Luo)和亮叶离蕊茶(*C. nitidissima* Chi)聚为一支,多齿红山茶、滇山茶和南山茶聚为一支,短柱茶[*C. brevistyla* (Hayata) Coh. St]、

小果油茶(*C. meiocarpa* Hu)和茶梅(*C. sasanqua* Thunb.)聚为一支,高州油茶(*C. gauchowensis* H. T. Chang)、油茶和大苞山茶(*C. granthamiana* Sealy)聚为一支,聚在一支的种类亲缘关系较近。红山茶组、油茶组(Sect. *Oleifera* H. T. Chang)和短柱茶组(Sect. *Paracamellia* Sealy)的种类聚在了不同的分支。



进化树分支上的数据为自展支持率,括号内编号为 GenBank 登录号 The data on the branches of evolutionary tree are bootstrap values, and the No. in brackets are the GenBank login numbers.

图3 基于叶绿体基因组的山茶科23个种类的系统发育树
Fig. 3 Phylogenetic tree of 23 species in Theaceae based on chloroplast genome

3 讨论和结论

通常情况下,高等植物叶绿体基因组的总GC含量在34%~40%之间,而且各部分分布不均匀^[7]。本研究结果显示:多齿红山茶叶绿体基因组的总GC含量以及大单拷贝区(LSC)、小单拷贝区(SSC)和反向重复区(IR)的GC含量与杜梨(*Pyrus betulifolia* Bunge)高度相似^[24],多齿红山茶叶绿体基因组中IR区的GC含量最高(42.98%),且明显高于SSC区,这是因为仅在IR区分布的4种rRNA基因具有较高的GC含量,而分布在SSC区的NADH脱氢酶基因的GC含量很低^[25]。

本文中,多齿红山茶叶绿体基因组的环状结构由1个LSC区、1个SSC区和2个IR区组成,与其他已报道的植物叶绿体基因组结构^[26]一致;多齿红山茶

有87个CDS基因、37个tRNA基因和8个rRNA基因,这与山茶属其他物种的分析结果相似,例如:‘龙井43’(*Camellia sinensis* ‘Longjing 43’)^[27]和油茶^[28]等;且多齿红山茶叶绿体基因组的结构和组成与典型被子植物苹果(*Malus pumila* Mill.)的叶绿体基因组^[29]相同,表明植物的叶绿体基因组具有高度的保守性。而对LSC、SSC和IR区边界结构进行比较后发现,IR和SSC区*ycf1*、*ndhF*和*trnN*基因的特异性导致不同物种间的叶绿体基因组的序列长度存在差异。通过自然选择,例如阴暗潮湿等环境因子,会使*ycf1*等调节光合作用的基因面临的选择压力不同,进而使得基因的进化速率不同^[30-31],这类进化速率不同的基因,可以组合起来作为DNA条形码来研究植物群落的系统发育研究^[32]。

本研究结果显示:在多齿红山茶叶绿体基因组中检测到的52个SSR位点中,75.0%的位点位于LSC

区,17.3%的位点位于 SSC 区,7.7%的位点位于 IR 区,不同区域中 SSR 位点的占比与其他山茶属植物的叶绿体基因组重复序列分析结果相似^[33],但多齿红山茶叶绿体基因组中重复序列的类型和位点的检出率与殷鑫等^[34]的结果不同,其原因是检索 SSR 位点的参数设定不同,导致输出的结果不同。当重复单元的范围设置较宽而长度下限较低时(单核苷酸到六核苷酸的最小重复值分别设置为 10、4、3、3、3、3),虽然位点的检出率很高,重复类型多,但会挖掘出较多难以检测的无效位点,后续实验中引物设计的成功率较低;反之当重复单元的范围设置较窄而长度下限较高时(单核苷酸到六核苷酸的最小重复值分别设置为 10、6、5、5、5、5),虽然位点的检出率很低、重复类型单一,但是引物设计成功率会相对更高,挖掘结果更有效。因此,若无特殊要求,重复单元长度的限定应该根据研究需要进行适当的调整^[35]。

本文中,多齿红山茶的相对同义密码子使用度(RSCU)大于 1.00 的密码子共有 30 个,其中 27 个以 A 或 U 结尾,3 个以 C 或 G 结尾,这一结果与大多数被子植物类似,都偏好使用 A 或 U 结尾的密码子^[36]。以同一基因或基因组为对象,当某物种的某一密码子 RSCU 值大于 1.60 时,为高使用频率密码子,其使用偏好性强于拟南芥 [*Arabidopsis thaliana* (Linn.) Heynh.], RSCU 值小于 0.60 时为低使用频率密码子,其使用偏好性弱于拟南芥^[37],本文中多齿红山茶高频率使用 AGA 编码 Arg,低频率使用 CUG、CAC、CGC、AAC、GCG 和 GAC 密码子分别编码 Leu、His、Arg、Asn、Ala 和 Asp。

聚类分析结果显示:小果油茶、短柱茶和茶梅聚为一支,从分子系统发育的角度支持闵天禄^[3]将短柱茶并入油茶组的观点。目前,有超过 32 个金花茶组(Sect. *Chrysantha* Chang)的分类群被发表,但是其相关分类依据仅基于叶表皮特征和若干 DNA 片段^[38,39]。本研究中,簇蕊金花茶(*Camellia fascicularis* H. T. Chang)仅与小花金花茶(*Camellia micrantha* S. Y. Liang et Y. C. Zhong)聚为一支,而金花茶、龙州金花茶和亮叶离蕊茶聚为另一支,从分子系统发育的角度支持李凤英等^[39]的基于叶表皮特征的聚类结果,簇蕊金花茶有别于金花茶组其他物种,小花金花茶与簇蕊金花茶的亲缘关系较近。此外,红山茶组物种被短柱茶组和油茶组种类交错间隔开,可能是由于在晚第三纪以来,古气候的变迁和亚洲山体的隆升等巨大

的环境变化导致了红山茶组和油茶组在新的环境中产生了进一步的分化和杂交^[40]。

综上所述,多齿红山茶叶绿体基因组长度为 156 778 bp,总 CG 含量为 37.33%,由 1 个 LSC 区、1 个 SSC 区以及 2 个 IR 区组成,对应 GC 含量分别为 35.34%、30.60%和 42.98%;共注释到 132 个基因,包括 87 个 CDS 基因(偏好使用以 A 或 U 结尾的密码子编码蛋白,高频率使用 AGA 编码 Arg)、37 个 tRNA 基因和 8 个 rRNA 基因,其中,*ycf1*、*ndhF* 和 *trnN* 基因的特异性是导致 IR 区边界产生差异的原因。检测到的 52 个 SSR 位点可作为分子标记位点进一步用于系统发育分析。本研究系统发育分析结果支持将短柱茶并入油茶组的观点。

参考文献:

- [1] 张宏达. 山茶属植物的系统研究[M]. 广州:中山大学学报编辑部, 1981.
- [2] 袁昌选, 杨 芹, 邓伦秀, 等. 长毛红山茶物种在天柱自然分布特征及人工繁殖技术研究[J]. 种子, 2019, 38(2): 41-45.
- [3] 闵天禄. 山茶属的系统大纲[J]. 云南植物研究, 1999, 21(2): 149-159.
- [4] HUANG H, SHI C, LIU Y, et al. Thirteen *Camellia* chloroplast genome sequences determined by high-throughput sequencing: genome structure and phylogenetic relationships [J]. BMC Evolutionary Biology, 2014, 14: 151.
- [5] 张韵洁, 李德铨. 叶绿体系统发育基因组学的研究进展[J]. 植物分类与资源学报, 2011, 33(4): 365-375.
- [6] OHYAMA K, FUKUZAWA H, KOHCHI T, et al. Chloroplast gene organization deduced from complete sequence of liverwort *Marchantia polymorpha* chloroplast DNA [J]. Nature, 1986, 322: 572-574.
- [7] 邢 钰, 冯慧喆. 叶绿体基因组的比较分析及系统发育研究[J]. 乡村科技, 2021(24): 82-83.
- [8] 李文哲. 植物叶绿体 DNA 与线粒体 DNA 的遗传变异[J]. 细胞生物学杂志, 1988, 10(1): 5-10.
- [9] 王章群, 解增言, 蔡应繁, 等. 系统发育基因组学研究进展[J]. 遗传, 2014, 36(7): 669-678.
- [10] 罗 纯, 张青林, 罗正荣. 第二代测序技术在植物遗传研究中的应用[J]. 广东农业科学, 2015(3): 186-192.
- [11] 陈昆松, 李 方, 徐昌杰, 等. 改良 CTAB 法用于多年生植物组织基因组 DNA 的大量提取[J]. 遗传, 2004, 26(4): 529-531.
- [12] JIN J J, YU W B, YANG J B, et al. GetOrganelle: a fast and versatile toolkit for accurate *de novo* assembly of organelle genomes [J]. Genome Biology, 2020, 21: 241.
- [13] SHI L, CHEN H, JIANG M, et al. CPGAVAS2, an integrated plastome sequence annotator and analyzer [J]. Nucleic Acids

- Research, 2019, 47: W65–W73.
- [14] ZHENG S, POCZAI P, HYVÖNEN J, et al. Chloroplast: an online program for the versatile plotting of organelle genomes[J]. *Frontiers in Genetics*, 2020, 11: 576124.
- [15] BEIER S, THIEL T, MÜNCH T, et al. MISA-web: a web server for microsatellite prediction[J]. *Bioinformatics*, 2017, 33(16): 2583–2585.
- [16] 张震, 许彦明, 陈永忠, 等. 油茶转录组测序与SSR特征分析[J]. *西南林业大学学报*, 2018, 38(6): 63–68.
- [17] MAREUIL F, DOPPELT-AZEROUAL O, MÉNAGER H. A public Galaxy platform at Pasteur used as an execution engine for web services[J]. *F1000Research*, 2017, 6: 1030.
- [18] 王占军, 吴子琦, 王朝霞, 等. 3个茶树品种 *WOX* 基因家族的进化及密码子偏好性比较[J]. *南京林业大学学报(自然科学版)*, 2022, 46(2): 71–80.
- [19] KATO H, STANDLEY D M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability[J]. *Molecular Biology and Evolution*, 2013, 30(4): 772–780.
- [20] KUMAR S, STECHER G, LI M, et al. MEGA X: molecular evolutionary genetics analysis across computing platforms[J]. *Molecular Biology and Evolution*, 2018, 35(6): 1547–1549.
- [21] KALYAANAMOORTHY S, MINH B Q, WONG T K F, et al. ModelFinder: fast model selection for accurate phylogenetic estimates[J]. *Nature methods*, 2017, 14(6): 587–589.
- [22] NGUYEN L T, SCHMIDT H A, VON HAESELER A, et al. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies[J]. *Molecular Biology and Evolution*, 2014, 32(1): 268–274.
- [23] LETUNIC I, BORK P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation[J]. *Nucleic Acids Research*, 2021, 49: W293–W296.
- [24] 李泳潭, 张军, 黄亚丽, 等. 杜梨叶绿体基因组分析[J]. *园艺学报*, 2020, 47(6): 1021–1032.
- [25] RAMAN G, PARK K T, KIM J H, et al. Characteristics of the completed chloroplast genome sequence of *Xanthium spinosum*: comparative analyses, identification of mutational hotspots and phylogenetic implications[J]. *BMC Genomics*, 2020, 21: 855.
- [26] 杨芮. 三个树莓类资源叶绿体基因组比较分析[D]. 雅安: 四川农业大学园艺学院, 2019: 22.
- [27] 陈春梅, 马春雷, 马建强, 等. 茶树 cpDNA 测序及基于 cpDNA 序列的山茶属植物亲缘关系研究[J]. *茶叶科学*, 2014, 34(4): 371–380.
- [28] 周安佩, 陈曦颜, 梁坚坤. 油茶叶绿体基因组特征及其系统分类[J]. *九江学院学报(自然科学版)*, 2021(1): 5–8, 14.
- [29] 金桂花, 陈斯云, 伊廷双, 等. 苹果叶绿体基因组特征分析[J]. *植物分类与资源学报*, 2014, 36(4): 468–484.
- [30] YANG X F, WANG Y T, CHEN S T, et al. PBR1 selectively controls biogenesis of photosynthetic complexes by modulating translation of the large chloroplast gene *Yefl* in *Arabidopsis*[J]. *Cell Discovery*, 2016, 2: 16003.
- [31] VITTI J J, GROSSMAN S R, SABETI P C. Detecting natural selection in genomic data[J]. *Annual Review of Genetics*, 2013, 47: 97–120.
- [32] 裴男才. 利用植物 DNA 条形码构建亚热带森林群落系统发育关系: 以鼎湖山样地为例[J]. *植物分类与资源学报*, 2012, 34(3): 263–270.
- [33] 李倩, 郭其强, 高超, 等. 贵州威宁红花油茶的叶绿体基因组特征分析[J]. *园艺学报*, 2020, 47(4): 779–787.
- [34] 殷鑫, 温强, 王建文, 等. 山茶属叶绿体全基因组微卫星特征分析及标记开发[J]. *分子植物育种*, 2018, 16(20): 6761–6769.
- [35] 王希, 陈丽, 赵春雷. 利用 MISA 工具对不同类型序列进行 SSR 标记位点挖掘的探讨[J]. *中国农学通报*, 2016, 32(10): 150–156.
- [36] 杨亚蒙, 焦健, 樊秀彩, 等. 桑叶葡萄叶绿体基因组及其特征分析[J]. *园艺学报*, 2019, 46(4): 635–648.
- [37] MUKHOPADHYAY P, BASAK S, GHOSH T C. Nature of selective constraints on synonymous codon usage of rice differs in GC-poor and GC-rich genes[J]. *Gene*, 2007, 400(1/2): 71–81.
- [38] 方伟, 杨俊波, 杨世雄, 等. 基于叶绿体四个 DNA 片段联合分析探讨山茶属长柄山茶组、金花茶组和超长柄茶组的系统位置与亲缘关系[J]. 2010, 32(1): 1–13.
- [39] 李凤英, 王玉国, 唐绍清. 山茶属金花茶组金花茶系的叶表皮特征及分类学意义[J]. *广西师范大学学报(自然科学版)*, 2001, 19(4): 75–79.
- [40] 闵天禄, 张文驹. 山茶属植物的进化与分布[J]. *云南植物研究*, 1996, 18(1): 1–13.

(责任编辑: 郭严冬)